社会互动视角下人际公平形成的脑机制*

张如倩 [2]　刘洁琼 [2]　李先春 [1,2,3]

（[1] 东南大学儿童发展与学习科学教育部重点实验室，南京 210096）

（[2] 华东师范大学心理与认知科学学院，上海 200062）

（[3] 上海长宁-华东师大精神卫生中心，上海 200335）

摘　要　最后通牒博弈任务被广泛用以探究公平行为，以往研究大多集中于对博弈中某一方决策行为和神经机制的探讨，但是人际公平可能是互动双方重复博弈的结果。因此只考察单个大脑活动，并不足以揭示由互动双方共同完成的社会认知活动的脑机制。因此，本研究结合修改版的最后通牒博弈任务和基于 fNIRS 的超扫描技术，从群体脑水平上考察人际公平形成的脑机制。行为结果显示，相比无惩罚条件，惩罚下提议者的分配金额更高，且惩罚力度越强，分配越趋近公平分配。fNIRS 的结果显示，惩罚下右侧背外侧前额叶皮层、顶下小叶和颞-顶联合区的脑间活动同步性显著强于无惩罚条件，而且两条件的分配金额差异越大，右侧顶下小叶的脑间活动同步性差异也越大。综上，脑间活动同步性可以作为惩罚下人际公平形成的客观脑指标，研究为探讨人际公平的内在机制提供了新的视角。

关键词　人际公平；惩罚；脑间活动同步性；功能近红外光谱技术；超扫描技术

1　引言

从古至今，“不患寡而患不均”的思想始终促使着人们不断追求公平（吴燕，周晓林，2012）。公平作为一种基本的社会准则，在人类生活中起着重要的作用，它不仅保障了个体的基本权利，同时也维护了社会的稳定发展。大量研究显示，当个体面对不公平时，个体的行为反应具有跨文化的一致性(Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006)，具体表现为牺牲自我利益以惩罚该种行为，而惩罚被证明是制裁违反公平规范的有效方法，有利于人际公平的形成(Fehr & Gächter, 2000; Henrich et al., 2006; Makwana, Grön, Fehr, & Hare, 2015; Wu et al., 2014)。

收稿日期：2018-08-10

*东南大学基本科研业务费资助项目（CDLS-2018-02），上海市卫计委重点专科项目（ZK2015B01），上海市卫计委项目（201540114）资助。

通讯作者：李先春，E-mail: xcli@psy.ecnu.edu.cn

目前对人际公平的研究主要采用行为经济学的博弈范式，其中最后通牒博弈任务（Ultimatum Game，UG）(Güth, Schmittberger, & Schwarze, 1982)是应用较为广泛的一种实验范式(Feng, Luo, & Krueger, 2015; Gabay, Radua, Kempton, & Mehta, 2014)。经典的最后通牒博弈任务包含两个角色，提议者（proposer）和接受者（recipient）。要求他们对一笔金钱进行分配，先由提议者进行分配，然后接受者决定是否接受当前的分配方案。如果接受者认可该分配方案，就按照该提议进行分配；如果接受者拒绝，则双方均无任何收益。Spitzer, Fischbacher, Herrnberger, Grön 和 Fehr(2007)采用修订版的最后通牒博弈任务探究了惩罚对于公平规范遵从的作用。惩罚条件下，若接受者认为分配方案不公平，可以采取惩罚措施，即接受者每拿出 1 元，就相应减少分配者 5 元。研究结果显示，无惩罚下提议者的分配金额仅为总金额的 10%，而惩罚下分配金额达到总金额的 40%，更接近平均分配（50%）。另外，惩罚条件下，分配者的分配越不公平（与 50 的差值越大），接受者的惩罚金额越大。由此可见，惩罚在促进人际公平方面起着重要作用(Makwana et al., 2015)。

随着脑成像技术的日益成熟，研究者们利用功能性磁共振成像技术（fMRI）深入探究最后通牒博弈任务中提议者和接受者遵从公平规范的神经机制。已有研究显示，相比无惩罚条件，惩罚条件下提议者的分配金额更高，分配过程中双侧背外侧前额叶皮层（DLPFC）、腹外侧前额叶皮层（VLPFC）、眶额皮层（OFC）以及双侧尾状核（caudate nucleus）的激活程度增强，惩罚条件下右侧背外侧前额叶皮层、左侧眶额皮层、右侧尾状核激活程度的增强与公平规范遵从程度的提升（惩罚下分配金额减去无惩罚下分配金额）呈正相关(Spitzer et al., 2007)。这表明遵守公平规范可以激活与认知控制(Miller & Cohen, 2001)、优势反应（自私自利）抑制(Aron, Robbins, & Poldrack, 2004)、惩罚威胁评估(O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001)以及不确定奖赏与惩罚刺激加工相关的脑区(Delgado, Locke, Stenger, & Fiez, 2003)。同时，研究者还发现公平规范遵从的神经通路具有偏侧化的特点，使用经颅直流电刺激（tDCS）增强右侧 DLPFC 的激活程度可显著提升提议者的分配金额，使其更接近平均分配；而抑制该脑区则会降低分配方案的公平性(Ruff, Ugazio, & Fehr, 2013)。

对接受者遵从公平规范脑机制的研究显示，存在两个交互作用的神经系统支持这一行为，包括反射性直觉系统（reflexive and intuitive system）和反省深思系统（reflective and deliberate system）。前者由前脑岛（AI）、杏仁核（amygdala）和腹内侧前额叶皮层（VMPFC）组成，涉及识别及评价社会规范违反行为，产生惩罚动机。后者由背侧前扣带回（dACC）、背外侧前额叶皮层（DLPFC）、腹外侧前额叶皮层（VLPFC）和背内侧前额叶皮层（DMPFC）组成，参与重评及冲突的调节，抑制自利行为，执行惩罚措施(Feng et al., 2015)。研究者们

同样采用了无创的干预技术探究了接受者在遵从公平规范中 DLPFC 功能的偏侧化，发现当右侧 DLPFC 被抑制时，接受者面对不公平分配时惩罚行为显著减少(Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Knoch et al., 2008; Knoch et al., 2006)。以上研究结果均提示右侧 DLPFC 在经济博弈中可能负责自利反应的抑制，让个体更好地遵从公平规范，一定程度上也暗示了公平规范遵从下提议者和接受者右侧 DLPFC 神经活动的相似性。

过往研究主要集中于对经济博弈中某一方决策行为和神经机制的探讨，但是人际公平的形成可能是互动双方重复博弈的结果(Spitzer et al., 2007)。而重复的博弈是一种互为因果的过程，个体的行为结果既是一种反馈，也是其互动对象行为的前提(Shaw et al., 2018)。因此只考察单个个体的大脑活动，并不足以揭示由互动双方共同完成的社会认知活动的脑机制。所以本研究中我们从单脑（single-brain）转向多脑（multi-brain）(Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012)，通过分析脑-脑互动的变化规律，从群体脑水平上揭示人际公平形成的神经机制。超扫描技术（hyperscanning）为人际互动下跨脑机制的探究提供了新途径(Montague et al., 2002)，通过对互动的多个脑活动进行同步记录，并分析脑间活动同步性（Interpersonal Neural Synchronization，INS），从群体角度上提供了社会互动脑机制的新机制。该技术已在社会认知神经科学的多个领域有所应用，研究涉及合作竞争(Cui, Bryant, & Reiss, 2012; Pan, Cheng, Zhang, Li, & Hu, 2017)、行为模仿(Holper, Scholkmann, & Wolf, 2012; Pan, Novembre, Song, Li, & Hu, 2018)、师生互动(Zheng et al., 2018)等。这些研究一致表明，社会认知相关脑区，特别是与“心理理论”相关脑区的 INS 是衡量人际互动重要的神经指标。在一项金钱分配游戏中，相比隔板遮挡的情境，面对面互动下双方右侧颞-顶联合区（TPJ）的 INS 增强。这可能是因为面对面的交流提供了外显的社会线索，例如面部表情、肢体动作等，使得双方能够更好地揣测对方心理，达成共享意图(Tang et al., 2016)。Liu, Saito 和 Oi(2015)采用轮流下棋的游戏探究了合作与竞争的神经机制。任务中包含两个角色，创建者（builder）和搭档（partner）。合作任务中，搭档需要辅助创建者复制给定的棋局。竞争任务中，搭档要尽可能阻碍创建者对棋局的复制。双方轮流摆放棋子，每人摆放四次。研究者发现，竞争任务中右侧额下回（IFG）出现了 INS，但是这一现象并不存在合作任务中。研究者认为竞争任务中，双方的卷入度更高，更加积极地推断他人的心理意图并调整自己的行为策略。由此可见，互动双方在相互博弈的过程中，“心理理论”相关脑区的交互性更强。

相比其他的脑功能成像技术，功能近红外光谱技术（functional near-infrared spectroscopy，fNIRS）在成本、对头动的容忍性以及研究的生态效度等方面具有优势(Scholkmann, Holper, Wolf, & Wolf, 2013)。因此，本研究将结合修改版的最后通牒博弈任务与基于 fNIRS 的超扫

描技术，设置惩罚和无惩罚两个条件，分析不同条件下分配金额和脑间活动同步性的差异，从互动视角下揭示人际公平形成的脑机制。基于以往研究结果，惩罚可以促使个体在经济决策中遵从公平规范，我们预期（1）惩罚下分配金额高于无惩罚条件，且更趋近平均分配。与此同时，遵从公平规范需要个体抑制优势反应（自利行为），且该过程中个体间的交互性更强。因此，我们预期（2）相比无惩罚条件，惩罚下提议者和接受者右侧背外侧前额叶以及“心理理论”相关脑区的 INS 更强。

2　方法

2.1　被试

共招募大学在读本科生 44 名（*M* = 21.3 岁，*SD* = 1.6 岁），其中男生 26 人，女生 18 人，共形成 22 对同性别被试对，每组被试在实验前互不认识。所有被试均为右利手，无脑部疾病或者精神疾病，视力或者矫正视力均正常。实验前，所有被试签署了知情同意。实验后，根据被试在实验过程中的实际表现给与相应的被试费。

2.2　实验任务

实验采用 Spitzer 等人(2007)基于最后通牒博弈任务修改而成的金钱分配任务，要求两名被试完成多轮金钱分配，金额为 100 元。实验一共分为两个条件，惩罚条件和无惩罚条件。惩罚条件下，由提议者（A）提出分配方案。如果接受者（B）认可分配方案，就按照该方案进行金钱分配。如果 B 认为分配方案不合理，可以通过减少自己拥有的金钱来对提议者进行惩罚，惩罚比率为 1:5，即接受者每拿出 1 元，就相应地减少提议者 5 元。因此，为了收益最大化，A 应适当调整分配方案，找到 B 所能接受的最小分配金额并努力维持；而 B 应通过惩罚的反馈方式，找到 A 所能给与的最大分配金额并尽力保持。无惩罚条件下，无论 B 是否认可 A 的分配方案，都只能无条件接受。

2.3　实验流程

2 名被试进入实验室后，面对面而坐，中间有隔板遮挡（如图 1A 所示）。实验开始前，被试双方通过抽签的方式来决定任务角色（提议者 A 和接受者 B），任务中角色固定。每一轮试次中，两人都先获得 25 元，这保证了惩罚条件下，即便 A 分配给 B 0 元，B 也有资本对 A 进行惩罚。之后，A 对 100 元进行分配。试次始于 3~6s 的注视点，接着呈现线索 2s，5:1 代表惩罚条件，0:0 代表无惩罚条件。任务线索呈现完毕后，A 思考分配方案 2s，而后提出分配方案，通过按“←”、“→”键控制绿色光标的移动，从而增加或者减少分配金额，时间不限，分配完后 A 按“↓”键进行确认。而后呈现该分配方案 2s，之后 B 进行反馈。在 5:1

的线索下，B 既可以接受，也可以通过减少 25 元中的部分或全部金钱来减少 A 相应的 5 倍金钱。通过按“1”、“3”键控制绿色光标的移动，B 可以增加或减少惩罚金额，惩罚完后按“2”键进行确认。需要注意的是，惩罚过程中，接受者仅能减少提议者的金额直至为 0 元。在 0:0 的线索下，B 只能按“2”键接受 A 的分配方案。最后向被试双方呈现分配结果 2s。任务中 A 和 B 电脑屏幕上所呈现的内容一致，即 B 可以实时看见 A 的分配情况，A 也能实时获取 B 的反馈情况（如图 1B 所示）。

任务共包含 50 个试次，惩罚条件和无惩罚条件各 25 个试次，两种条件的试次以伪随机的方式进行。任务开始前，被试静息 40s。完成 25 个试次后，被试休息 2 分钟，然后完成剩余的 25 个试次。实验前告知被试，被试费由基本被试费（完成实验）和任务过程中积累金额的一定百分比组成。

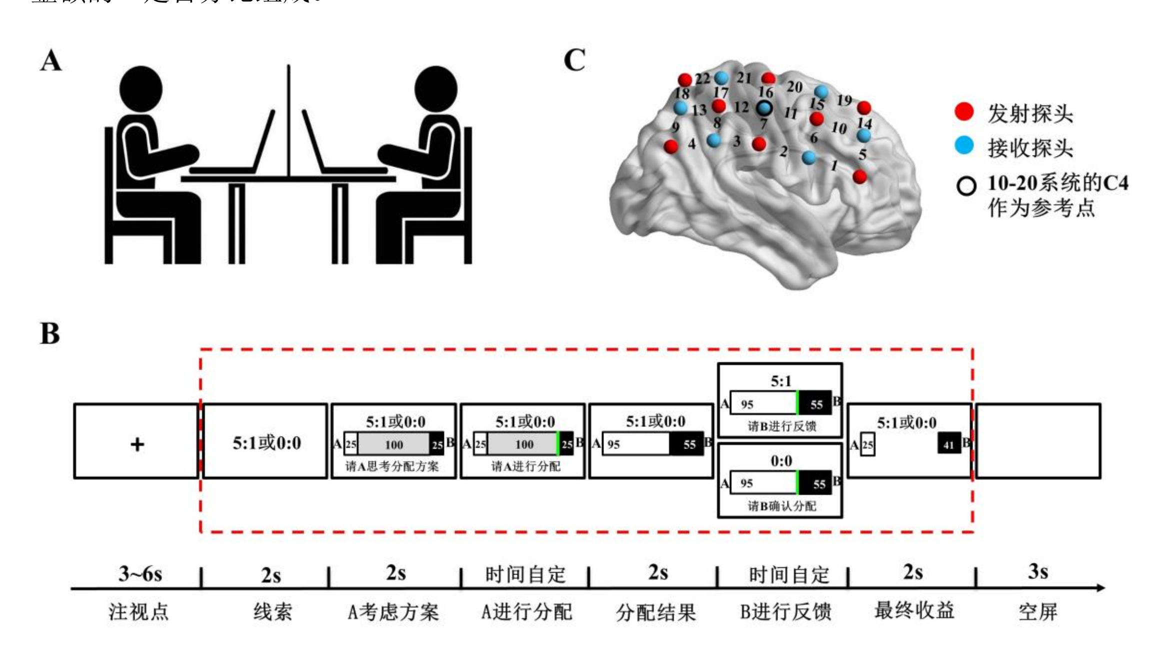


图 1　实验场景、实验流程和探头板定位

2.4　数据收集

通过 E-prime 2.0 软件呈现任务刺激并采集行为学数据。另外，采用 Hitachi（日立）ETG-4000 来记录被试在实验任务中大脑氧合血红蛋白（oxy-hemoglobin，Hbo）以及脱氧血红蛋白（deoxy-hemoglobin，Hbr）浓度的变化。将一块 3×5 的探头板置于泳帽中，放置在每名被试的大脑右侧。根据国际标准 10-20 系统，将探头板正中间的探测器放置在 C4 的位置，探头板沿着矢状参考曲线进行放置（如图 1C 所示），覆盖脑区为额叶和顶叶。每个探头板由 8 个发射器和 7 个探测器组成，探头之间间隔 3cm，每个探头板有 22 个通道，每个通道的具体位置参考 Jichi University 提供的定位模板（http://www.jichi.ac.jp/brainlab/virtual_registration/Result3x5_E.html，如表 1 所示）(Reindl,

Gerloff, Scharke, & Konrad, 2018)。采样频率为 10Hz。

表 1　通道位置

通道	MNI 坐标			AAL 分区		布鲁德曼分区	
	x	y	z	脑区	概率	脑区	概率
1	59.49	21.37	20.43	右侧三角部额下回	0.67	44-岛盖部，布洛卡区部分	0.53
2	66.87	-7.18	29.10	右侧中央后回	1.00	43-中央下区	1.00
3	67.53	-33.40	34.90	右侧缘上回	1.00	40-缘上回，威尔尼克区部分	0.72
4	59.42	-59.03	33.70	右侧角回	1.00	39-角回，威尔尼克区部分	0.97
5	50.69	36.73	27.05	右侧额中回	0.52	45-布洛卡区三角区	1.00
6	58.80	7.96	36.56	右侧中央前回	0.97	6-前运动皮层和辅助运动皮层	0.78
7	63.88	-20.28	43.97	右侧缘上回	0.76	1-初级躯体感觉皮质	0.83
8	60.37	-45.14	46.84	右侧顶下小叶	0.60	40-缘上回，威尔尼克区部分	1.00
9	48.36	-68.79	43.83	右侧角回	1.00	39-角回，威尔尼克区部分	1.00
10	47.43	24.55	43.41	右侧额中回	0.99	44-岛盖部，布洛卡区部分	0.51
11	54.49	-4.25	50.69	右侧中央前回	0.55	6-前运动皮层和辅助运动皮层	0.97
12	56.89	-31.30	54.81	右侧顶下小叶	0.72	1-初级躯体感觉皮质	0.68
13	49.68	-55.49	55.37	右侧顶下小叶	0.84	40-缘上回，威尔尼克区部分	0.81
14	33.83	37.74	46.08	右侧额中回	0.98	9-背外侧前额叶皮层	1.00
15	42.48	13.34	56.74	右侧额中回	1.00	9-背外侧前额叶皮层	0.79
16	47.64	-17.53	62.57	右侧中央前回	0.52	4-初级运动皮层	0.65
17	43.34	-42.84	63.31	右侧顶上小叶	0.55	2-初级躯体感觉皮质	0.52
18	32.38	-63.83	61.62	右侧顶上小叶	0.99	7-躯体感觉联合皮层	1.00
19	27.55	24.62	58.16	右侧额上回	0.78	8-前额眼动区	0.99
20	34.22	-1.43	65.56	右侧额上回	0.66	6-前运动皮层和辅助运动皮层	1.00
21	35.98	-29.29	70.73	右侧中央后回	0.64	4-初级运动皮层	0.84
22	28.73	-53.67	70.81	右侧顶上小叶	1.00	7-躯体感觉联合皮层	1.00

2.5　数据分析

2.5.1　行为数据

计算惩罚和无惩罚条件下每一位提议者 A 对接受者 B 的平均分配金额、A 和 B 不同条件下的平均收益（包括分配前的初始资金 25 元），并使用配对样本 *t* 检验比较不同实验条件下分配金额的差异以及 A 和 B 平均收益的差异。*t* 检验的效应量为 Cohen's *d*，0.2、0.5、0.8 分别是小、中、大效应量。另外，计算惩罚条件下接受者 B 对提议者 A 不公平分配（分配

金额小于 50 元）的惩罚力度：$\frac{5\times B\text{对}A\text{的惩罚金额}}{50\times\text{不公平试次数量}-A\text{对}B\text{不公平分配的总金额}}$（分子：不公平试次（即分配金额小于 50 元）的总惩罚金额；分母：不公平试次中提议者亏欠接受者的总金额。比值越大，说明惩罚力度越大），并使用皮尔逊相关计算惩罚力度与惩罚条件下分配金额的相关性。

2.5.2 fNIRS 数据

在 Matlab（2014a）的操作环境下，使用 NIRS_SPM 对数据进行预处理。采用血液动力响应函数（Hemodynamic Response Function，HRF）和离散余弦变换，截止频率 128s(Discrete Cosine Transform，DCT)进行低通滤波和高通滤波，从而消除仪器产生的噪音和呼吸心跳等引起的生理噪音(Ye, Tak, Jang, Jung, & Jang, 2009)。

鉴于氧合血红蛋白对任务刺激更加敏感(Hoshi, 2003)，在本研究中，仅将其纳入进一步的数据分析。使用小波变换相干（Wavelet Transform Coherence, WTC）计算任务期间（线索阶段到反馈阶段）的脑间活动同步性（Interpersonal Neural Synchronization，INS）(Cui et al., 2012; Grinsted, Moore, & Jevrejeva, 2004)。首先，对时间序列 *x(n)*进行连续小波变换（Continuous Wavelet Transform，CWT）：$W_n^X(s)=\sqrt{\frac{\delta t}{s}}\sum_{n'=1}^{N}x_{n'}\,\psi_0[(n'-n)\frac{\delta t}{s}]$（n：时间点，s：小波尺度， δt：采样周期（例如 0.1s），N：时间序列长度）。然后，对两个时间序列 *x(n)*和 *y(n)*进行交叉小波变换（Cross Wavelet Transform，XWT）：$W^{XY}(n,s)=W^X(n,s)W^{Y*}(n,s)$,*为复共轭。最后，计算两个时间序列的小波变换相干性，$R_n^2(s)=\frac{|S(s^{-1}W_n^{XY}(s))|^2}{|S(s^{-1}W_n^X(s))|^2\cdot|S(s^{-1}W_n^Y(s))|^2}$，S 为平滑算子。

为了确保所选频段与任务相关，基于以往超扫描研究(Nozawa, Sasaki, Sakaki, Yokoyama, & Kawashima, 2016; Pan et al., 2018; Zheng et al., 2018)，将 0.01Hz~0.2Hz（5s~100s）中每一个频段的 INS 进行平均，该频段可以避免由生理信号，例如心跳（~1Hz）、呼吸（0.2~0.3Hz）(Kamran & Hong, 2014; Pierro, Hallacoglu, Sassaroli, Kainerstorfer, & Fantini, 2014)造成的脑间活动同步性。将平均以后的值进行 fisherz 转换。对每一个频段下 22 个通道的 INS 分别进行配对样本 *t* 检验，$p < 0.005$。结果显示，惩罚条件下 0.13~0.2Hz、0.03~0.04Hz 的 INS 显著高于无惩罚条件，但无惩罚条件下，没有频段的 INS 高于惩罚条件（如图 2 所示）。因此，选取 0.13~0.2Hz、0.03~0.04Hz 为研究频段。使用配对样本 *t* 检验分别考察这两个频段下每一个通道 INS 条件间的差异，对 *p* 值进行 FDR 校正(Benjamini & Hochberg, 1995)，并生成 *t* 值热图，通过 xjview 工具箱（http://www.alivelearn.net/xjview8/）和 BrainNet Viewer 工具

箱(Xia, Wang, & He, 2013)（http://www.nitrc.org/projects/bnv/）将热图至于 3D 的脑模型上。

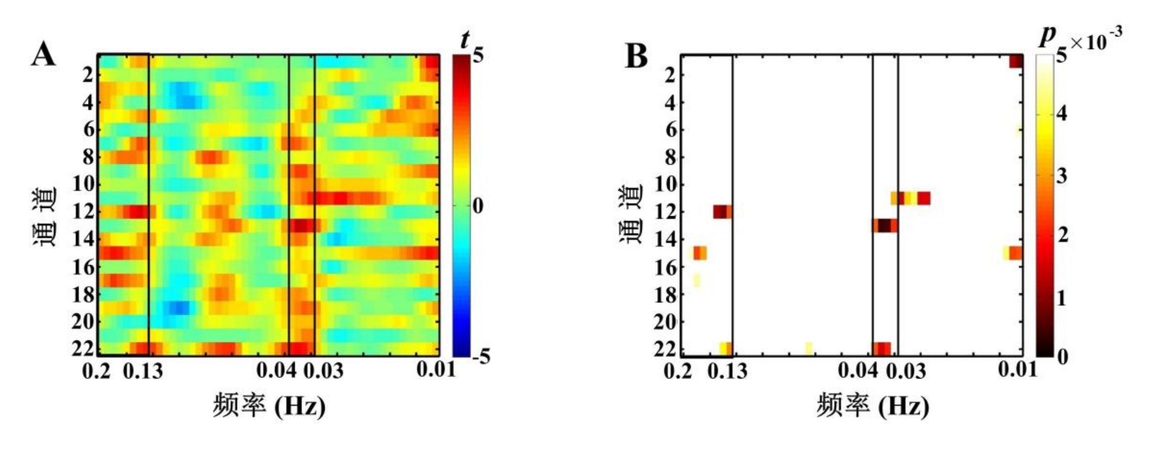


图 2　频段的选取

3　结果

3.1　行为结果

22 对被试的互动情况如图 3 所示。对提议者不同实验条件下的分配金额进行配对样本 *t* 检验，发现惩罚条件下提议者的分配金额（41.39 ± 9.28）高于无惩罚条件（30.27 ± 17），$t(21) = 3.91$, $p < 0.001$, Cohen's $d = 0.83$, 95%CI: 5.21~17.03（如图 4A 所示）。

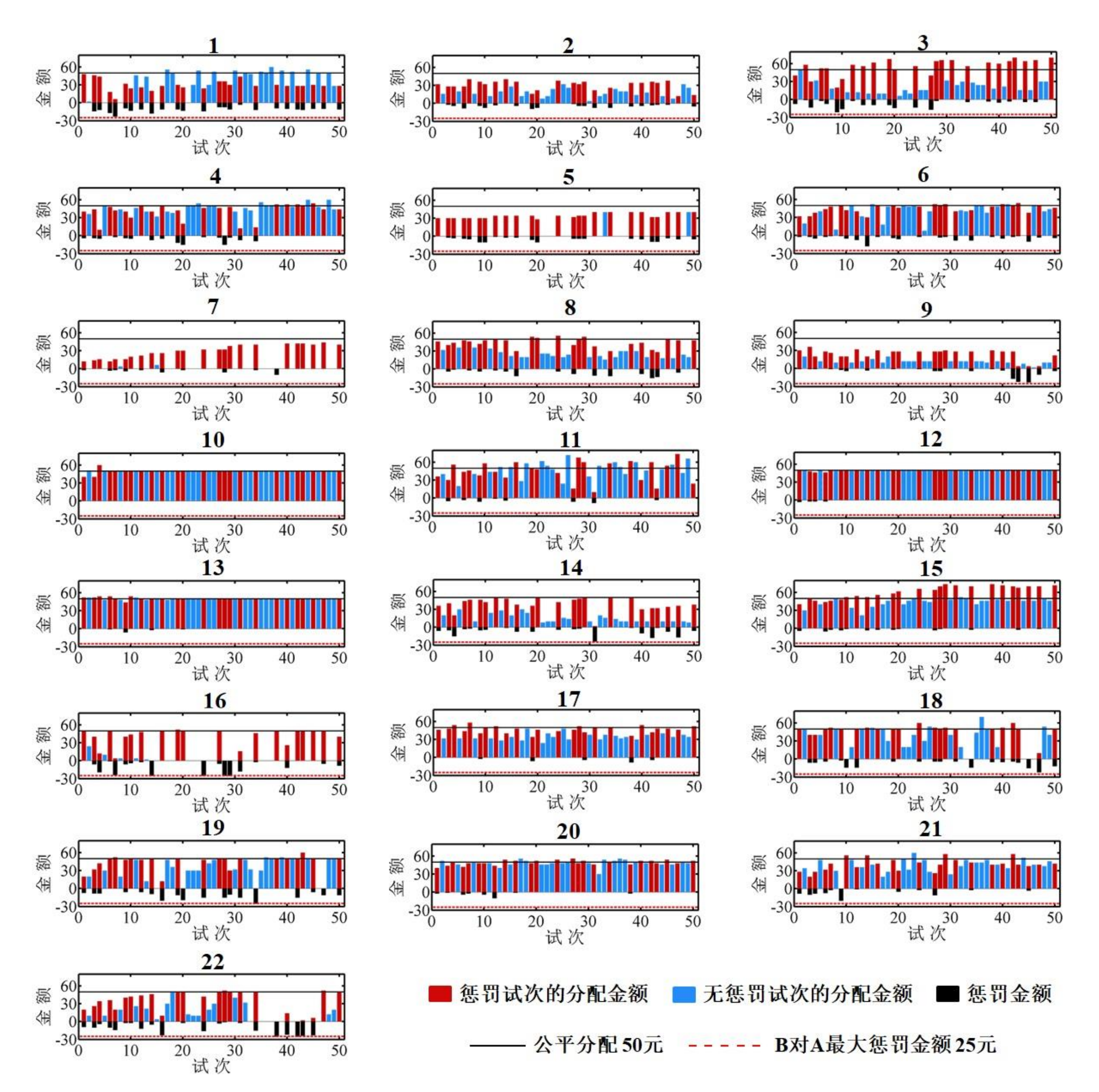


图 3　22 对被试的互动情况

惩罚力度和惩罚下平均分配金额的相关分析的结果显示，惩罚力度越大，惩罚条件下提议者的分配金额越高，r = 0.43，p = 0.04（如图 4B 所示）。另外，对不同实验条件下提议者和接受者的平均获益进行配对样本 t 检验，发现惩罚条件下，提议者（62.76 ± 14.53）和接受者（62.22 ± 11.18）的平均获益没有显著差异，$t(21)$ = 0.15, p = 0.88, Cohen's d = 0.03, 95%CI: -7~8.07；而无惩罚条件下，提议者的平均获益（94.72 ± 17）显著高于接受者（55.27 ± 17），$t(21)$ = 5.44, p < 0.01, 95%CI: 24.38~54.53（如图 4C 所示）。

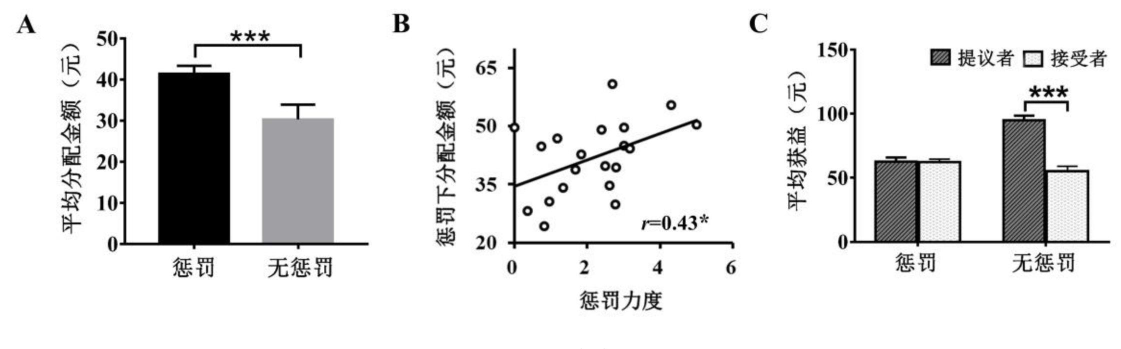


图 4　行为结果

注：图 A：不同实验条件下的平均分配金额；图 B：惩罚条件下惩罚力度与分配金额的相关性；图 C：不同实验条件下提议者和接受者的平均获益金额

3.2　fNIRS 结果

分别考察 0.13~0.2Hz 和 0.03~0.04Hz 频段脑间活动同步性条件间的差异。结果显示，当频段为 0.13~0.2Hz 时，通道 12 和 15 惩罚条件下的 INS 显著高于无惩罚条件，通道 12: $t(21) = 3.56$, $p = 0.04$, Cohen's $d = 0.76$, 95%CI: 0.01~0.05; 通道 15: $t(21) = 3.45$, $p = 0.03$, Cohen's $d = 0.74$, 95%CI: 0.01~0.05，p 为 FDR 校正后（如图 5A&B 所示）。斯皮尔曼等级相关显示，通道 12 的惩罚条件与无惩罚条件分配金额的差值与 INS 的差值呈正相关，$\rho = 0.45$，$p = 0.04$（如图 5C 所示）。

当频段为 0.03~0.04Hz 时，通道 9、13 和 22 惩罚条件下的 INS 显著高于无惩罚条件，通道 9: $t(21) = 2.99$, $p = 0.05$, Cohen's $d = 0.64$, 95%CI: 0.01~0.05; 通道 13: $t(21) = 4.21$, $p = 0.01$, Cohen's $d = 0.90$, 95%CI: 0.02~0.05; 通道 22: $t(21) = 3.56$, $p = 0.02$, Cohen's $d = 0.76$, 95%CI: 0.01~0.05，p 为 FDR 校正后（如图 5D&4E 所示）。

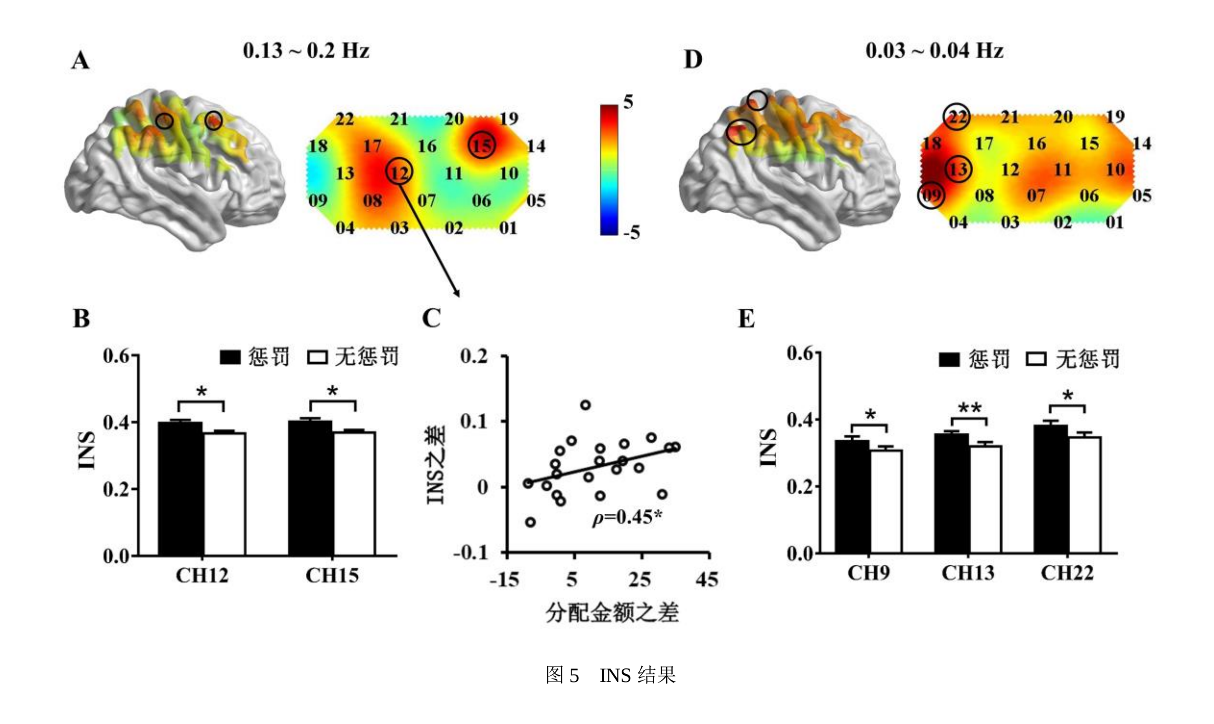


图 5　INS 结果

4　讨论

本研究采用真实的人-人互动情境下的最后通牒博弈任务，结合基于 fNIRS 的超扫描技术，从群体互动视角下探究了人际公平形成的脑机制。行为结果显示，惩罚下个体的分配金额更高，且更接近平均分配。面对不公平分配，接受者的惩罚力度越大，提议者的分配越趋向平均分配。fNIRS 的结果显示，惩罚下提议者和接受者右侧背外侧前额叶皮层和“心理理论”相关脑区的脑间活动同步性（INS）增强，且顶下小叶的脑间活动同步性的增强与分配金额的差异（惩罚-无惩罚）具有显著的相关关系。

研究中设置了两个实验条件，惩罚条件和无惩罚条件。惩罚下，接受者面对不公平分配时可以采取惩罚措施以纠正提议者的违规行为；无惩罚下，接受者没有惩罚权利，只能接受提议者的任何分配方案。研究结果验证了惩罚对于促进提议者遵从公平规范的有效性和必要性。但与前人研究结果(Spitzer et al., 2007)相比，无惩罚下提议者的分配金额提高了 20 元。这可能是因为本研究中提议者和接受者在惩罚和无惩罚条件下进行了多轮分配，导致两个实验条件相互有影响，使得提议者的分配行为也受到了潜在的惩罚威胁的影响。我们还发现接受者面对不公平分配时的惩罚力度越大，对促进提议者遵守公平规范的效应越明显。社会公平的形成一方面需要个体自觉遵从行为规范，另一方面对违规行为也应做出合理的惩罚。如果接受者利益短视，没有适度地对违规行为进行惩罚，就会导致提议者为利益最大化，保持不公分配，甚至继续降低分配金额以试探接受者的底线。另外，当不公分配已经发生时，惩

罚也是消除个体间差异的一种方式。研究结果显示，惩罚下提议者和接受者最终的平均收益没有显著的差异。可能的原因有两点，一是人们普遍厌恶不公平的分配，其中既包括劣势不公平厌恶（自己比别人获得的少）,也包含优势不公平厌恶（自己比别人获得多）(Fehr & Schmidt, 1999; Gao et al., 2018)。二是适度的惩罚对于提议者已经具有威慑作用，接受者也需尽可能减少惩罚金额以求收益最大化。

fNIRS 的结果显示，惩罚下提议者和接受者右侧背外侧前额叶皮层（DLPFC，通道 15）、右侧顶下小叶（IPL，通道 12 和 13）和右侧颞-顶联合区（TPJ，通道 9）的脑间活动同步性显著增强。以往研究显示，右侧 DLPFC 在经济博弈中与认知控制、目标维持相关(Miller & Cohen, 2001; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003)。在最后通牒博弈任务中，提议者做出公平分配以及接受者拒绝不公平分配时，右侧背外侧前额叶均会显著激活(Knoch, Gianotti, Baumgartner, & Fehr, 2010; Weiland, Hewig, Hecht, Mussel, & Miltner, 2012)。在惩罚条件下，为避免因惩罚遭受损失，提议者会趋向公平分配；而接受者从长远利益出发，会“自损八百”以促使提议者遵守公平规范。在这样的经济博弈中，互动双方需要抑制优势反应以获得最大收益。由此惩罚下提议者和接受者右侧背外侧前额叶皮层神经活动的一致性也反映了这一共同的心理现象。

在最后通牒任务中，提议者做出分配的时候往往会进行策略性的思考（strategic considerations），对当前分配可能导致的接受者的情绪和行为反应进行估计(Weiland et al., 2012)。同样，当接受者拥有惩罚权利来抑制不公平行为时，也需对惩罚后提议者的行为反应加以推断。这种对他人心理状态进行解释和预测的认知活动称之为心理理论（mentalizing）(Corbetta, Patel, & Shulman, 2008; Gallagher & Frith, 2003)。本研究中，我们发现惩罚下与“心理理论”相关脑区的 INS 显著增强，包括右侧顶下小叶（IPL）和颞-顶联合区（TPJ）。IPL 是镜像神经元系统的组成部分，研究发现镜像神经元系统是社会认知重要的神经机制，可以促进个体自我加工与社会理解(Casartelli & Molteni, 2014; Dapretto et al., 2006; Marsh & Hamilton, 2011)。右侧 TPJ 也是心理理论加工的重要脑区，大量的研究证据表明右侧 TPJ 不仅与推断他人信念有关(Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006)，还涉及表征他人行为意图、整合意图与行为结果等(Koster-Hale, Saxe, Dungan, & Young, 2013; Young & Saxe, 2008)。“心理理论”相关的 INS 广泛存在于各种形式的社会互动中，例如联合注意(Bilek et al., 2015)、同步唱歌(Osaka et al., 2015)、教学过程中的师生互动(Zheng et al., 2018)、人际交谈(Dai et al., 2018; Jiang et al., 2012)等，且 INS 越强，个体间的合作或互动越好。因此，INS 可以作为衡量人际互动水平和共享意图的客观指标，而惩罚下顶下小叶 INS 的增量与分

配金额增量的正相关可能反应了惩罚下人际公平的提高伴随着博弈中互动双方更强的心理交互。

本研究首次尝试采用超扫描技术，从社会互动视角下探究人际公平形成的神经机制，并得出了一些结果，但是研究仍存在不足之处。第一，研究发现，有些个体无论是惩罚条件还是无惩罚条件始终都能遵从公平规范；而有些个体无惩罚条件下几乎不做任何分配，这可能是个体之间自利水平的不同导致的。因此，未来研究在探究人际公平问题的时候，可以将人格特质纳入考察。第二，遵从公平规范是一个复杂的社会认知过程，不仅涉及大脑皮层脑区，也与皮层下脑组织（例如脑岛、杏仁核等）紧密相关，但是功能近红外光谱技术无法对皮层下的神经信号进行探测，未来研究者们可以采纳高密度 EEG 源成像技术(Seeber et al., 2019)或者 fMRI 技术，在保证研究生态效度的同时，更全面地揭示人际公平形成的脑-脑机制。

5 结论

本研究首次结合修改版的最后通牒博弈任务和超扫描技术探讨了社会互动中人际公平形成的脑-脑机制，本研究发现惩罚促进人际公平形成与互动双方前额叶、颞-顶联合区等部位更强的脑间活动同步有关。

参考文献

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*(4), 170–177.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468–1474.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300.

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., . . . Meyer-Lindenberg, A. (2015). Information flow between interacting human brains: Identification, validation, and relationship to social expertise. *Proceedings of the National Academy of Sciences, 112*(16), 5207–5212.

Casartelli, L., & Molteni, M. (2014). Where there is a goal, there is a way: What, why and how the parieto-frontal mirror network can mediate imitative behaviours. *Neuroscience & Biobehavioral Reviews, 47*, 177–193.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron, 58*(3), 306–324.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage, 59*(3), 2430–2437.

Dai, B., Chen, C., Long, Y., Zheng, L., Zhao, H., Bai, X., . . . Lu, C. (2018). Neural mechanisms for selectively tuning in to the target speaker in a naturalistic noisy situation. *Nature Communications, 9*(1), 2405.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience, 9*(1), 28–30.

Delgado, M. R., Locke, H. M., Stenger, V. A., & Fiez, J. A. (2003). Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 27–38.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Feng, C., Luo, Y. J., & Krueger, F. (2015). Neural signatures of fairness-related normative decision making in the ultimatum game: A coordinate-based meta-analysis. *Human Brain Mapping, 36*(2), 591–602.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367–388.

Gabay, A. S., Radua, J., Kempton, M. J., & Mehta, M. A. (2014). The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 47*, 549–558.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends in Cognitive Sciences, 7*(2), 77–83.

Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous-and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences, 115*(33), E7680–E7689.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics, 11*(5/6), 561–566.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences, 16*(2), 114–121.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767–1770.

Holper, L., Scholkmann, F., & Wolf, M. (2012). Between-brain connectivity during imitation measured by fNIRS. *Neuroimage, 63*(1), 212–222.

Hoshi, Y. (2003). Functional near-infrared optical imaging: Utility and limitations in human brain mapping. *Psychophysiology, 40*(4), 511–520.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience, 32*(45), 16064–16069.

Kamran, M. A., & Hong, K. S. (2014). Reduction of physiological effects in fNIRS waveforms for efficient brain-state decoding. *Neuroscience Letters, 580*, 130–136.

Knoch, D., Gianotti, L. R., Baumgartner, T., & Fehr, E. (2010). A neural marker of costly punishment behavior. *Psychological Science, 21*(3), 337–342.

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cerebral Cortex, 18*(9), 1987–1990.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*(5800), 829–832.

Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. L. (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences, 110*(14), 5648–5653.

Liu, T., Saito, H., & Oi, M. (2015). Role of the right inferior frontal gyrus in turn-based cooperation and competition: A near-infrared spectroscopy study. *Brain and Cognition, 99*, 17–23.

Makwana, A., Grön, G., Fehr, E., & Hare, T. A. (2015). A neural mechanism of strategic social choice under sanction-induced norm compliance. *eNeuro, 2*(3), ENEURO.0066-14.2015.

Marsh, L. E., & Hamilton, A. F. D. C. (2011). Dissociation of mirroring and mentalising systems in autism. *Neuroimage, 56*(3), 1511–1519.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167–202.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., . . . Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *Neuroimage, 16*(4), 1159–1164.

Nozawa, T., Sasaki, Y., Sakaki, K., Yokoyama, R., & Kawashima, R. (2016). Interpersonal frontopolar neural synchronization in group communication: An exploration toward fNIRS hyperscanning of natural interactions. *Neuroimage, 133*, 484–497.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*(1), 95–102.

Osaka, N., Minamoto, T., Yaoi, K., Azuma, M., Shimada, Y. M., & Osaka, M. (2015). How two brains make one synchronized mind in the inferior frontal cortex: fNIRS-based hyperscanning during cooperative singing. *Frontiers in Psychology, 6*, 1811.

Pan, Y., Cheng, X., Zhang, Z., Li, X., & Hu, Y. (2017). Cooperation in lovers: An fNIRS-based hyperscanning study. *Human Brain Mapping, 38*(2), 831–841.

Pan, Y., Novembre, G., Song, B., Li, X., & Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage, 183*, 280–290.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience, 1*(3-4), 245–258.

Pierro, M. L., Hallacoglu, B., Sassaroli, A., Kainerstorfer, J. M., & Fantini, S. (2014). Validation of a novel hemodynamic model for coherent hemodynamics spectroscopy (CHS) and functional brain studies with fNIRS and fMRI. *Neuroimage, 85*, 222–233.

Reindl, V., Gerloff, C., Scharke, W., & Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *Neuroimage, 178*, 493–502.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342*(6157), 482–484.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755–1758.

Scholkmann, F., Holper, L., Wolf, U., & Wolf, M. (2013). A new methodical approach in neuroscience: Assessing inter-personal brain coupling using functional near-infrared imaging (fNIRI) hyperscanning. *Frontiers in Human Neuroscience, 7*, 813.

Seeber, M., Cantonas, L. M., Hoevels, M., Sesia, T., Visser-Vandewalle, V., & Michel, C. M. (2019). Subcortical electrophysiological activity is detectable with high-density EEG source imaging. *Nature Communications, 10*(1), 753.

Shaw, D. J., Czekóová, K., Staněk, R., Mareček, R., Urbánek, T., Špalek, J., . . . Brázdil, M. (2018). A dual-fMRI investigation of the iterated Ultimatum Game reveals that reciprocal behaviour is associated with neural alignment. *Scientific Reports, 8*(1), 10896.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56*(1), 185–196.

Tang, H., Mai, X., Wang, S., Zhu, C., Krueger, F., & Liu, C. (2016). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Social Cognitive and Affective Neuroscience, 11*(1), 23–32.

Weiland, S., Hewig, J., Hecht, H., Mussel, P., & Miltner, W. H. (2012). Neural correlates of fair behavior in interpersonal bargaining. *Social Neuroscience, 7*(5), 537–551.

Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., . . . Zhou, X. (2014). Neural basis of increased costly norm enforcement under adversity. *Social Cognitive and Affective Neuroscience, 9*(12), 1862–1871.

Wu, Y., & Zhou, X. L. (2012). The context-dependency of fairness processing: Evidence from ERP study. *Acta Psychologica Sinica, 44*(6), 797–806.

[吴燕, 周晓林. (2012). 公平加工的情境依赖性: 来自 ERP 的证据. *心理学报*, *44*(6), 797–806.]

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS One, 8*(7),

e68910.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *Neuroimage, 44*(2), 428–447.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage, 40*(4), 1912–1920.

Zheng, L., Chen, C., Liu, W., Long, Y., Zhao, H., Bai, X., . . . Lu, C. (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Human Brain Mapping, 39*(7), 3046–3057.

Neural mechanisms of fairness formation in the perspective of social interactions

ZHANG Ruqian[2] LIU Jieqiong[2] LI Xianchun[1,2,3]

([1] *Key Laboratory of Child Development and Learning Science （Southeast University）, Ministry of Education, Nanjing 210096, China*)

([2] *The School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China*)

([3] *Shanghai Changning-ECNU Mental Health Center, Shanghai 200335, China*)

Abstract

Interpersonal fairness plays an important role in human life. Punishment is one of the most fundamental factors in the formation of interpersonal fairness. Although many neuroscience studies have used interactive economic games to explore the neural mechanisms that underlie fair norms' enforcement and compliance, single brain studies cannot make researchers achieve a complete understanding of the intrinsic mechanisms related to dynamic interactions. Hyperscanning techniques can allow researchers to measure the brain activity of two or more persons simultaneously while they complete cognitive tasks under natural conditions. Therefore, we used fNIRS-based hyperscanning to evaluate changes in behavior and interpersonal neural synchronization (INS) during the formation of interpersonal fairness.

The experimental paradigm referred to the adapted ultimatum game (UG). There were two conditions in our study, punishment and voluntary condition. In the punishment condition, the proposer (A) introduced a distribution scheme. When the recipient (B) accepted it, both of them

received money according to the distribution scheme. If B thought that this distribution was unfair, he or she could punish A by spending all or part of their money to reduce A's gains. One RMB invested in punishment led to a reduction in A's gains by 5 RMB. However, in the voluntary condition, B was obligated to accept the offer no matter how much A allocated to him or her. The task included 50 trials altogether, 25 trials for each condition. The trials of two different conditions were presented in a pseudo-random way. A total of 44 university students participated in this study, including 13 male and 9 female pairs. Before the experiment, pairs of participants drew lots to decide their roles.

Behavioral results showed that compared with the voluntary condition, the average transfer amount was significantly higher in the punishment condition and much closer to the equal split. Moreover, in the face of unfair distribution, more money was transferred to recipients when they punished more the proposers. Consequently, there was no significant difference between proposers' and recipients' final gains in the punishment condition. However, in the voluntary condition, proposers gained more. Results from fNIRS revealed that in the punishment condition, the INS of three brain areas was higher than that of the voluntary condition, including the dorsolateral prefrontal cortex (channel 15), inferior parietal lobule (channel 12,13), and temporo-parietal junction (channel 9). Additionally, the transfer difference (punishment minus control) was positively associated with an increase of INS in the inferior parietal lobule.

These results suggest that punishment can promote the formation of interpersonal fairness, accompanied by enhanced interpersonal neural synchronizations in the brain areas related to strategic decision-making and theory of mind.

Key words fairness; punishment; interpersonal neural synchronization; fNIRS; hyperscanning.